Language Learning & Technology
ISSN 1094-3501

October 2017, Volume 21, Issue 3
pp. 131–152

ARTICLE

Enhancing extensive reading with data-driven learning

Gregory Hadley, *Niigata University*

Maggie Charles, *Oxford University Language Centre*

Abstract

This paper investigates using data-driven learning (DDL) as a means of stimulating greater lexicogrammatical knowledge and reading speed among lower proficiency learners in an extensive reading program. For 16 weekly 90-minute sessions, an experimental group (12 students) used DDL materials created from a corpus developed from the Oxford Bookworms Graded Readers, while a control group (10 students) had no DDL input. Both classes were required to read a minimum of 200,000 words during the course. An embedded-experiment design (Edmonds & Kennedy, 2017) was adopted consisting of both qualitative and quantitative forms of investigation. Quantitative data from the Vocabulary Levels Test by Nation and Beglar (2007) and a C-test (Klein-Braley & Raatz, 1984) constructed from an upper-level Bookworms reader found statistically significant lexicogrammatical improvements for both groups, but greater improvement took place within the control group. Qualitative data derived from a repertory grid analysis of student constructs revealed several possible reasons for the experimental group's lack of engagement with DDL. The study concludes that careful attention to students' learning preferences and a softening of the DDL approach may ensure better results with lower proficiency learners.

Keywords: *Data-Driven Learning, Extensive Reading, Lower Proficiency Learners*

Language(s) Learned in this Study: *English*

APA Citation: Hadley, G., & Charles, M. (2017). Enhancing extensive reading with data-driven learning. *Language Learning & Technology*, *21*(3), 131–152. Retrieved from http://llt.msu.edu/issues/october2017/hadleycharles.pdf

Introduction

Data-driven learning (DDL) is a student-centered inductive method of language learning, in which learners explore grammar and vocabulary issues using a corpus. A corpus is a large database of language that is searched with software called a concordancer, which typically provides key words in context (KWIC). Students learn through repeated exposure to authentic occurrences of the same lexical items or phrases (see Figure 1).

The classic form of DDL advocates direct access to corpus data. Learners investigate language questions on their own, and develop a strong sense of autonomy (Johns, 1994, p. 296). Studies have found that DDL results in significant improvement among advanced and intermediate level language learners (e.g., Boulton & Cobb, 2017; Braun, 2007; Gordani, 2013; Granger, Hung, & Petch-Tyson, 2002; Hirata & Hirata, 2015; Sun & Wang, 2003). However, studies on the impact of DDL at the novice or intermediate-low level (e.g., Boulton, 2009; Hadley, 2002; Mizumoto & Chujo, 2015; St. John, 2001) have been mixed, due in part to the linguistic challenges presented by the corpus data itself.

This article reports on the first stage of a 4-year funded research project exploring the applications of DDL with lower proficiency learners—which in the East Asian context, refers to students who have succeeded in attaining higher than average scores on standardized language tests, but whose proficiency in class tends to be either inconsistent or uneven in terms of the relative balance between receptive and productive

language skills. Building upon the findings of earlier research, we link DDL to extensive reading (ER) and investigate, using both statistical and qualitative measures, whether a comprehensible form of DDL can enhance the reading speed, vocabulary recognition, and grammatical improvement of lower proficiency students. First, however, we will consider a number of issues underpinning the rationale of this study.

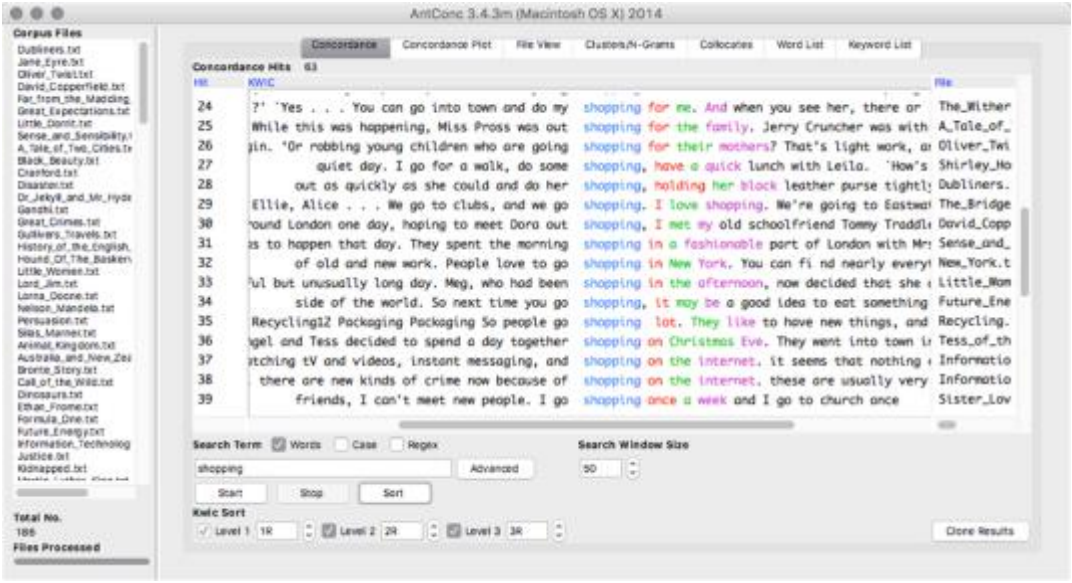


Figure 1. Example of corpus data using AntConc concordance software (Anthony, 2014)

Review and Research Questions

Starting with nascent ideas from the late 1960s (McEnery & Wilson, 1997, p. 12) and culminating with the work of Johns during the 1990s (Johns, 1991), DDL was initially dismissed by many as "the fringe activity of a few eccentrics" (Leech, 1997, p. 4). Today, however, "a substantial number of teachers are using corpora in their teaching" (Tribble, 2015, p. 55). The spread of DDL has resulted in it becoming more nuanced, so that it now operates within a spectrum of *soft* to *hard* instructional forms (see Bernardini, 2004; Gabrielatos, 2005; Gilquin & Granger, 2010). Hard DDL represents Johns' (1991) classic approach, where the corpus consists of authentic texts and the students work individually in accessing the corpus and deducing answers from the data. Soft DDL, on the other hand, typically refers to teachers preselecting corpus data and using it in a purposeful way to raise student consciousness about specific lexicogrammatical items within the target language.

While DDL has been found to positively impact the language development of learners with higher levels of proficiency, its results among lower proficiency learners have been mixed—in part because some lower proficiency learners find the KWIC presentation of lexical items to be distractingly decontextualized (see also Charles, 2007), and because considerable levels of scaffolding are needed for training learners in how to do DDL at lower levels (Hadley, 2001, 2002). A compounding difficulty is the linguistic challenge of corpora that are compiled from English sources that greatly exceed the lower proficiency level. Writing on the issue of linguistic difficulty, Chambers (2010) concludes that corpus linguists have not sufficiently heeded the input hypothesis (Krashen, 1989), the zone of proximal development (Vygotsky, 1978), or the cognitive load theory (Sweller, 1994)—all of which state that learnable second language input must be challenging, but comprehensible and integrated within a wide range of learning activities (Doughty, 2003; Loschky, 1994; Tomlinson, 2016).

One rich, but largely overlooked, source of comprehensible input can be found in ER. First used within the context of second language learning by Palmer (Kelly, 1969, p. 139), ER has evolved to become an "approach to the teaching and learning of reading in which learners read large quantities of material that is

within their linguistic competence" (Grabe & Stoller, 2013, p. 286). Dao's (2014) summary of several years of ER research provides convincing evidence for ER's capacity to improve many aspects of learners' proficiency including grammatical accuracy, vocabulary comprehension, reading comprehension, and speed—though others rightly note that more research is still needed on the degree to which learning rates and vocabulary retention are affected (Waring & Takaki, 2003).

Proponents of ER have traditionally stressed that reading should be for pleasure, that students should have access to a wide variety of graded readers, and that the use of dictionaries, chalk-and-talk teacher instruction, and testing should be scrupulously avoided (Day & Bamford, 2002, 2004). Nevertheless, questions are arising today as to whether the traditionally inductive approach of ER is the most efficient way for students to learn (e.g., Bieri, 2015; Waring & McLean, 2015). Grabe and Stoller (2013, p. 99) are representative of those calling for more research into how ER can be improved through other forms of instruction and learning resources such as dictionaries or word lists.

The literature on DDL has been largely silent on whether it can be used to aid learners in vocabulary retention while reading for meaning. A notable exception, however, can be found in Allan (2009), who created a small corpus of graded readers in order to provide comprehensible language input for her intermediate learners. Though limited in terms of time, corpus size, and focus, Allan's pioneering study provided evidence to suggest that DDL used with graded readers helped students to quickly acquire a number of lexical chunks. More significantly, she found that the simplified language in graded readers did not pose a threat to authenticity, because the lexical chunks in the graded readers were no different from those in texts created by native speakers for native consumption. "For the teacher looking for a way into DDL with lower-level learners," she concludes, "it seems that graded corpora may offer a reasonable balance of accessibility and authenticity in the data it provides" (p. 30).

Our research continues the work started by Allan (2009) and takes up the challenge of transcending a strictly inductive approach to ER. By combining the elements of DDL, lower proficiency learners, comprehensible input, and calls for ER enhancement within a study of longer and broader reach, we hope to provide further empirical evidence that would "not just show that DDL works, but in what conditions" (Boulton, 2007, p. 14). This article reports on the first stages of this investigation by quantitatively studying test results and qualitatively investigating student affective responses to a semi-hard form of DDL. The qualitative aspect of this study is especially important because the tendency of many Japanese institutions is in coaxing, rather than compelling, learners to study. As a result, it has become increasingly necessary for teachers to discern what appeals to learners and craft their lessons accordingly. The research questions were formulated as follows:

1. Does the use of comprehensible DDL in an ER program for lower proficiency learners lead to greater gains in reading speed, lexical knowledge, and grammatical improvement than achieved by comparable students in a non-DDL ER program?
2. What are some possible affective concerns among lower proficiency learners that could influence the potential effectiveness of DDL?

Methods

Setting

Our investigation took place at a Japanese national university in the spring semester of 2015, and was carried out within the university's Applied English Course—an elective subject that encourages the use of innovative instructional methods. Student levels were mixed and it was not known in advance who would attend. Moreover, the course relied upon student interest; participants normally devoted time to similar courses only after satisfying the demands of required courses. Two classes participated in this study during the same semester, and met once a week for 90 minutes over 16 weeks.

Materials

We approached Oxford University Press (OUP) for permission to use their Bookworms Graded Readers as corpus source material. This series consisted of 192 books divided into seven levels of increasing lengths and difficulty. The starter level featured short readings averaging around 1,400 words, often presented in a graphic novel type of format, while level six contained short versions of classic novels, with up to 2,500 headwords and over 30,000 words of text.

OUP agreed to our project proposal, and generously provided a full set of Oxford Bookworms for classroom use. Another set purchased by the university library's Self-Access Learning Center (SALC) further ensured that students would have ample opportunities for ER both in and out of class. For security purposes, OUP created the corpus on our behalf, and omitted a small number of books for either copyright or technical reasons. The resulting Bookworms Corpus contains 186 books from all seven levels with a total of 1,715,160 tokens (17,670 word types). Additional details regarding the corpus can be found in Table 1. In order to protect their proprietary rights, however, OUP stipulated that the corpus could be used only in class, and not placed on any electronic platform within the university's network. This posed special challenges, which will be discussed shortly.

Table 1. *Breakdown of Bookworms Corpus*

Stages	Number of Titles	Average Words per Book	Word Types	Tokens
Starter	28	1,412	1,727	40,794
Stage 1	49	5,753	5,932	292,424
Stage 2	41	6,508	6,146	275,424
Stage 3	31	10,395	7,375	327,090
Stage 4	22	16,327	7,765	363,691
Stage 5	9	24,361	5,706	224,944
Stage 6	6	31,195	6,066	191,086

Methods and Instruments

An embedded-experiment design—sometimes referred to as a concurrent, nested mixed-methods strategy—was adopted for this study. Edmonds and Kennedy (2017, pp. 189–191) explain this as a method used when exploring issues occurring within a real-world context, where both qualitative and quantitative data sources are available. The aim was to gather findings that could stimulate hypotheses on the efficacy of a particular treatment or intervention. In this case, the phenomenon was of lower proficiency learners studying English via ER, and the treatment was a semi-hard form of DDL. Embedded-experiment designs begin with a quantitative pre-test, and qualitative data is collected as the intervention is taking place. A quantitative post-test is administered, and follow-up qualitative analysis is used to further unpack the quantitative data. Rather than highly specific confirmation research or a more subjective case study, an evidence-based, mixed-methods approach of the type found in the embedded-experiment design facilitates a multi-layered understanding of what students learn, how they respond, and how they might best succeed in their studies (Hesse-Biber & Johnson, 2015). The qualitative and quantitative features of this investigation are shown in Table 2.

Table 2. *Methods and Instruments Used in this Investigation*

Research Question	Method	Instruments
Does the use of comprehensible DDL in an ER program for lower proficiency learners lead to greater gains in reading speed, lexical knowledge, and grammatical improvement than achieved by comparable students in a non-DDL ER program?	Quantitative	Pre- and post-test of Vocabulary Levels Test, C-tests, and speed reading tests using *t*-test measures of dependent and independent means Cohen's *d*
What are some possible affective concerns among lower proficiency learners that could influence the potential effectiveness of DDL?	Qualitative & Mixed Method	Classroom observation Portfolio analysis Personal Construct Repertory Grids

Quantitatively, pre-tests used *t*-tests of independent means to determine whether the control and experimental groups were the same or distinct statistical groups. Post-tests employed *t*-tests of dependent means to ascertain if each group had improved to the level of statistical significance during the course of the semester, and a *t*-test for independent means for investigating whether the DDL group had improved to the point where it had become distinct from the control group at a statistically significant level. Cohen's *d* was used to measure the size of effect in the data. The statistical analyses were carried out using StatPlus:mac Pro (AnalystSoft, 2016). Qualitatively, a classroom observation log, photographs of student classroom activity, and student portfolio analyses were employed during DDL use. An instrument known as Personal Construct Repertory Grids, which enabled research participants to share thoughts and opinions with minimal researcher interference, was used following the post-test.

Personal Construct Repertory Grids were developed over 50 years ago by Kelly, the founder of Personal Construct Psychology (Kelly, 1955, 1963). Fundamental to personal construct psychology are the notions of *elements*, *constructs*, and *bipolar constructs*. Elements are the "people, events, objects, ideas, institutions, and so on" (Cohen, Manion, & Morrison, 2003, p. 338), which are "well-known and personally meaningful" to research informants (Shaw, 1980, p. 10). Easterby-Smith (1981) explains that it is helpful "to think of elements as being the objects of people's thoughts, and constructs as the qualities that people attribute to these objects" (p. 11). Kelly (1963) writes that mental constructs are bi-polar and that in order to discover how someone understands value-laden constructs such as *sincere*, *cheerful*, or *refined*, the analyst needs to learn something about the contrasting limits within which these values are framed (pp. 105–108). Although there have been numerous modifications to Kelly's original procedure for conducting repertory grid interviews, all variations share the common goal of studying how individuals and groups see the world around them (Bannister & Fransella, 1986). In English language teaching (ELT), repertory grids have been used to discover the reasons why adults dropped out of an ESL class (Rowsell, 1992), as a means for exploring teachers' perceptions of language education (Roberts, 1999), and as a tool for understanding the cultural expectations for good teachers in Japan (Hadley & Evans, 2001).

Repertory grids first encourage informants to think about empirical issues specifically related to an area of research interest. Informants then go through several cycles of choosing a sample of these empirical events and affix value statements to them. A rating system is then used to link the empirical events to the value statements in a meaningful way. The resulting data can be analyzed in a number of ways, and a *cognitive snapshot* of informant notions can be developed (Figure 2).

The specific procedures used for grid elicitation came from Shaw and McKnight (1981) and Jankowicz (2004). Each student received a repertory grid large enough for five elements and bipolar constructs. As can be seen in Figure 2, they were provided with some elements and asked to write other language learning activities across the top of the grid. Students then looked at three randomly chosen elements and identified what two of them had in common and what differentiated them from the third. This produced a set of mental constructs, which were written on the sides of the grid. Students rated each activity on a scale from 1 to 5 according to how well it corresponded with each bi-polar construct. All of the students' repertory grids

were combined into one large grid, and analyzed using GridSuite 4 (Fromm & Bacher, 2014) and RepGrid V (Shaw & Gaines, 2010). These programs use principle components analysis to locate the most strongly rated constructs, and chart the relationship of elements to constructs as a cluster map. Repertory grids enabled us to gain insight into which activities were liked and disliked, and the reasons behind students' attitudes.

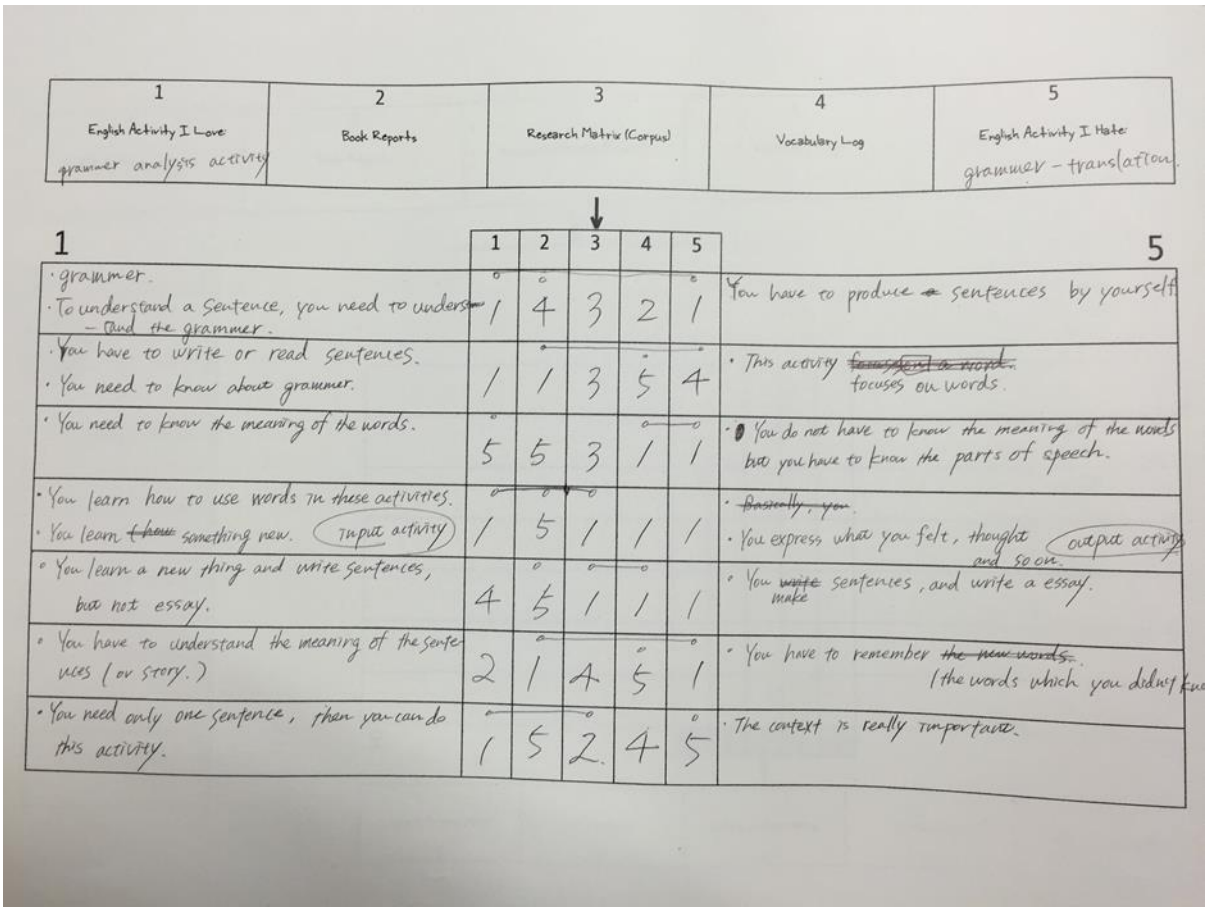


Figure 2. Example of a completed repertory grid

Participants

An unexpectedly large number of international students joined both classes. In both the experimental and control groups, Chinese learners formed the majority, followed by Japanese, French, and Korean students (see Table 3). This provided an excellent opportunity to conduct research on DDL with various nationalities, mirroring the environment in many English language centers around the world. For this reason, the class with the most diversity in terms of nationality and gender was chosen as the experimental group.

Table 3. *Research Participants*

	Gender		Nationality			
Class	**Male**	**Female**	**Chinese**	**Japanese**	**French**	**Korean**
Control Group (N = 10)	3	7	5	4	1	0
Experimental Group (N = 12)	4	8	5	4	2	1

In Japanese universities, nearly three quarters of ER classes are conducted with groups of 25 or fewer

learners (Shimamoto, 2008; Yoshida, 2014, p. 21), meaning that our group sizes were fairly average for our context. Admittedly, however, the number of participants in this study is a potentially problematic issue that will be addressed later.

Pre-Test Data Collection

A series of pre-tests were administered to both the experimental and control groups in order to ascertain their levels in reading speed, vocabulary knowledge, and grammatical knowledge. Both groups took a speed reading test (Quinn, Nation, & Millett, 2007) and the Vocabulary Levels Test (VLT) for the 2,000-, 3,000-, 5,000-, and 10,000-word levels from Nation's resource website (Nation & Beglar, 2007). The speed reading test contained 563 words and 10 question items, and the VLT featured 30 items for each of the four word levels. We opted for these tests as exploratory tools at the beginning of this project, because they are widely used and have been tested for reliability and validity (Coxhead, 2000; Schmitt, Schmitt, & Clapham, 2001). Furthermore, VLT scores have been shown to correlate with overall L2 proficiency (Milton, 2013), which provided us with useful background information on our participants.

Students also completed a 100-item partial deletion cloze, or C-test (Klein-Braley & Raatz, 1984), created from the Bookworms version of *David Copperfield* by Charles Dickens. C-tests are a modified form of cloze testing widely used to investigate a subject's knowledge of written discourse items such as context cohesion, syntax, and strategic textual comprehension (Bachman, 1982). They correlate highly with grammar tests and reading tests (Alderson, 1979), and have been found to be ideal for testing lexicogrammatical proficiency (Bowen, Madsen, & Hilferty, 1985, p. 376).

Post-Test Data Collection

Student attrition occurred at two per class and their data were not included in the study. The post-tests were conducted following the same procedure as the pre-tests. In addition, repertory grids were administered to students in the experimental class. Each individual completed a grid according to the procedure described earlier.

Instructional Procedures

The ER program commenced with an orientation class that discussed the practices and instructional procedures employed. ER, the maintaining of portfolios, and in-class expansion activities were common to both classes.

Extensive Reading

All students were required to read a minimum of 200,000 words over the semester. Both classes read from the same sets of hard-copy Oxford Bookworms Graded Readers. Their progress was tracked using the MReader site administered by Kyoto Sangyo University and the Extensive Reading Foundation. This provided short online quizzes for the books that students had read, and gave them credit for the word amounts of each book after successfully passing each quiz. Students were trained in how to access the site, create individual MReader accounts, and take the online tests. Throughout the semester, student progress was monitored and anonymous progress lists were regularly presented in order to encourage regular reading.

A selection of books from all levels of Oxford Bookworms Graded Readers was brought to each class, and 30 to 45 minutes of each class was devoted to silent sustained reading. This was to ensure that students had at least one time in the week when we knew they had been reading. The principal investigator (PI) taught the students the *five finger rule*, which helped students to choose graded readers where no more than two or three unknown words were encountered on a page. During sustained silent reading sessions, the PI observed book choices and reading habits in order to enhance in-class activities. Outside class, students had free access to a wide range of graded readers in the university SALC, but in both classes, students bonded with the Bookworms Graded Readers and read solely from that series.

Portfolios and Expansion Activities

All students kept a portfolio of vocabulary and grammatical items by using a vocabulary log worksheet (see Appendix A). Our vocabulary logs, which were modified versions of the vocabulary notebooks found in Schmitt and Schmitt (1995), required students to write down unknown words from their reading, and engage in short exercises, such as drawing a picture of the word and crafting their own sentence using the new item. Students in both classes also wrote one book report per week for a book they had read that week.

In class, following Eichhorst and Shearon (2013), the typical format for each 90-minute class began with 30 minutes devoted to an expansion activity. The control class would engage in various conversation activities of the type found in activity books designed for use with graded readers (e.g. Bamford & Day, 2004; Greenwood, 1988). The experimental class participated in these activities and also in DDL-based activities using the Bookworms Corpus to stimulate further vocabulary comprehension. While DDL activities had a prominent place in the instructional procedures within the experimental class, a balance needed to be struck between DDL and other activities, because the PI learned from earlier attempts at using DDL with beginners that learners would quickly become either bored or overwhelmed when working with corpus-based materials (Hadley, 2001). Approximately 30 minutes of each class focused on silent sustained reading, and classes ended with a speed-reading test from Nation's resource website. Students charted their progress at the end of each class, and their scores were recorded and monitored.

To prepare the learners in the experimental group for DDL, they were given an introduction to corpora and DDL using the classroom screen projector and a laptop that had the Bookworms Corpus and AntConc (Anthony, 2014) concordance software installed. Initially, we wished to investigate a purely hard DDL approach for this stage of the investigation, but the stipulations from OUP precluded allowing students unfettered direct access to the corpus materials. This necessitated our semi-hard approach, in that corpus material was printed and provided for students, but we still adopted Johns' (1991) procedure of *identify, classify, generalise* to structure the work, meaning that students were encouraged to interact with the concordance materials independently by finding patterns, categorizing them, and then developing hypotheses about the lexical items. Limiting class time for DDL meant that students in the experimental group were expected to spend a considerable time outside of class working with the corpus materials provided. This, we observed, resulted in a heavier workload than the control class.

Results

After a brief consideration of the pre-test findings, we summarize the qualitative observational data of both classes and our interventions, as the learners' response to DDL was instrumental in understanding both the quantitative post-test scores and the results of the repertory grid analysis.

Pre-Test

An independent-samples *t*-test was conducted to compare the language proficiency of the experimental and control groups. Language proficiency was assessed in three areas: lexico-grammatical proficiency via the C-test, speed reading, and lexical knowledge as measured by the VLT. Maximum scores were 100 points for the C-test, speed reading was measured in words per minute (wpm) through the rubric provided by Nation and Beglar (2007), and the highest possible score for each level of the VLT was 30 points.

No significant difference emerged from the C-test scores of the control class (M = 66.45, SD = 11) and experimental class (M = 64.23, SD = 12): $t(22) = -0.466$, $p = .645$; $d = -0.18$. No significant difference was found in the wpm scores for the control class (M = 134, SD = 19) and experimental class (M = 143, SD = 23), as assessed on the speed reading pre-test: $t(22) = 1.031$, $p = .313$; $d = 0.44$. However, the high standard deviations indicate the degree to which proficiency was mixed in these classes, and even though the effect sizes by some estimations appear to be limited (see Plonsky & Oswald, 2014) it seemed clear to us that some in the experimental class had higher levels of proficiency in this area than those in the control class. Scores for the experimental and control classes for all levels of the VLT test, however, revealed no

significant differences, suggesting that at the beginning of the study, both groups were essentially the same statistically (see Table 8). The VLT scores also show that the overall L2 proficiency of participants in this study was, on average, between the A2 and B1 levels (Milton, 2013) within the Common European framework of reference for languages (Council of Europe, 2001).

Observed Student Response

The affective response of the experimental group to the semi-hard DDL experience was disappointing. With the portfolio material, the performance of the experimental class was observed as less enthusiastic than the control class almost from the outset. While the average rate of return on weekly book reports over the entire semester was roughly similar for both classes (control class = 77%; experimental class = 74%), the quality of the experimental class reports was appreciably lower. This was determined not only by the physical appearance of the reports, but, with the exception of those written by the two French students, most of the experimental class reports were short and lacking in specificity. Many appeared to have been finished at the last minute.

Both classes submitted similar totals in terms of vocabulary log sheets (experimental class = 44; control class = 40), but here again, over half of the log sheets in the experimental class came from four female students (two Japanese and two Chinese). The rest of the students in the experimental class completed only one to four sheets each over the semester, and these were only filled out partially. Students in the control class were fewer in number, produced almost the same amount in total as the larger experimental class, and each produced an average of five sheets during the semester, but these were in most cases completely finished and work was maintained on a regular basis throughout the semester. It should be noted that while students had a deadline for turning in book reports on Fridays (when they knew the PI made a note of who had and had not submitted work), we waited until the end of the semester to check the vocabulary logs. The lack of constant surveillance allowed space for students to follow the dictates of their personal levels of motivation. However, the students in the experimental class largely skimped on doing their logs, while the control class worked diligently on them. We believe this was a result of the time commitment needed for working with the DDL materials.

With the DDL component, students were encouraged to ask questions, either in class or in writing. However, the PI received only one question per week, and this only after considerable coaxing. When the PI asked students about what they had learned from the concordance readouts, or if they had any questions, it was discovered that, with the notable exception of one Japanese female student, the rest of the students were doing little more than lightly reading over the sheets, folding them, and then slipping them away into a disorganized corner of the portfolio.

As an intervention, we further softened our semi-hard approach by creating a worksheet that we called the research matrix (see Appendix B). The intention here was to give students more structure in order to more explicitly guide them in the process of inductive learning, which is a major feature of DDL. In subsequent lessons, the research matrix helped them create hypotheses in class. Outside of class, they were asked to write samples gleaned from the corpus data, and then expand upon these to create their own sentences. Unfortunately, however, the matrix had the opposite effect in eliciting questions from the learners. With the exception of one female Japanese student and one male Chinese student, the experimental group ceased asking questions. Of the two who would ask questions, neither completed the out-of-class work on their research matrix sheets.

Given this lack of success, we next attempted to emulate the approach of the co-principal investigator, who used DDL in her classes at a UK university language center. We brought the corpus into the classroom, projected the concordance software on screen, and made it available for students to consult in class. The PI fielded questions from the students, and using the concordance software in class, they explored the answers together. A few students were willing to give this approach a chance, but for most, working on something that could be seen by everyone in the class was threatening, and many demonstrated an aversion to using the laptop for technical reasons. It became clear that the experimental class lacked commitment and

engagement with DDL, and had shifted towards what Schlecty (2011, p. 16), describes as "retreatism."

In both the experimental and control classes, students displayed more visible energy and enthusiasm when engaged in more traditional expansion activities. It was observed that the energy of the students in the experimental group picked up even more when they were given activities for sharing their personal insights, or when discussing what they had learned from their reading. Therefore, we created classroom discussion tasks: as homework, students were given some corpus data and a research matrix worksheet, which they were asked to complete before the next class. In the following class, students then formed small groups and taught each other what they had learned from their corpus data studies. This had a better effect, but the overall response at best reached only what Schlecty (2011, pp. 16–19) describes as "ritual compliance", in that the class did only the minimum necessary to avoid any possible negative consequences.

Possible reasons for the student response to DDL will be considered later in the qualitative findings of this project. First, however, let us consider whether semi-hard DDL had an appreciable effect on the learners' lexicogrammatical proficiency and reading rates.

Post-tests: Research Question 1

C-Test

C-test post-test results (see Table 4 and Table 5) indicate that, while both groups improved significantly after approximately 10 weeks of ER, the control group improved more, with means comparisons indicating that they became distinct from the experimental group (one-tail) at a statistically significant level, which was also confirmed by *p*-value analysis. The effect size in the independent means post-test analysis, was negligible, and the slight shift into negative territory suggests that, if this study had continued, the mean change might have gone in the opposite of the predicted direction. Regardless, on the C-test, the experimental class did not improve more than the control class.

Table 4. *C-Test Paired Post-Test Results with Dependent Means*

Post-test	*df*	*M*	*SD*	*t*	*t* Two-Tail (.05)	*t* Two-Tail (.01)	*p*	*d*	Conclusion
Control Class	9	82	8.87	5.661	2.262	3.25	< .001	1.62	H_1
Experimental Class	11	75	9.59	6.649	2.200	3.11	< .001	1.38	H_1

Table 5. *C-Test Combined Post-Test Results with Independent Means*

Post-test	*df*	*M*	*SD*	*t*	*t* One-Tail (.05)	*t* One-Tail (.01)	*p*	*d*	Conclusion
Control Class	20	82	8.87	1.826	1.725	2.846	0.041	-0.82	H_1
Experimental Class		75	9.59						

Speed Reading

Speed reading means were based on wpm levels measured on Nation's Speed Reading Test Key. Pre-test analysis suggested that the experimental class had more readers who were faster, with an average speed of 143 wpm. This was due, however, to a few quick readers in the experimental class, and a few exceptionally slow students in the control class, which is confirmed by the relatively small effect size. Apart from these students, *p*-levels suggest that acceptance of the null hypothesis was reasonable. Again, in both groups reading speed improved over the semester, with a significant effect. However, the control class here did better again, with double the effect (see Table 6 and Table 7).

Table 6. *Speed Reading Paired Post-Test Results with Dependent Means*

Post-test	*df*	*M*	*SD*	*t*	*t* Two-Tail (.05)	*t* Two-Tail (.01)	*p*	*d*	Conclusion
Control Class	9	174	18	6.203	2.262	3.25	< .001	2.06	H_1
Experimental Class	11	168	23	4.538	2.200	3.11	< .001	1.13	H_1

Table 7. *Speed Reading Combined Post-Test Results with Independent Means*

Post-test	*df*	*M*	*SD*	*t*	*t* One-Tail (.05)	*t* One-Tail (.01)	*p*	*d*	Conclusion
Control Class	20	174	18	-0.638	1.725	2.846	.041	-0.30	H_0
Experimental Class		168	23						

Lexical Measures

Lexical measures (see Table 8) were generated using the VLT, which investigated the probability of lexical knowledge at the 2,000-, 3,000-, 5,000-, and 10,000-word levels (Nation & Beglar, 2007). The results here were unremarkable, because the students had mastered most words at the 2,000 level, while conversely having very little knowledge of words at the 10,000-word level. It is no surprise that little movement took place. The control class showed significant improvement, however, in the 3,000- and 5,000-word levels.

Table 8. *Lexical Measures*

	VLR 2000		VLR 3000		VLR 5000		VLR 10,000	
	M	*t*	*M*	*t*	*M*	*t*	*M*	*t*
Pre-test (Independent)								
Control Class	29	-1.163	24	0.032	19	0.182	5	-0.197
Experimental Class	28		24		20		4	
Post-test (Paired)								
Control Class	29	1.176	26	**3.222**	22	**2.508**	5	0.348
Experimental Class	28	1.076	25	1.604	20	1.160	5	1.134
Post-test (Independent)								
Control Class	29	-1.179	26	-1.346	22	-0.991	5	-0.044
Experimental Class	28		25		20		5	

Note. t-Scores in bold indicate significance (p < .05)

Total Words Read

Word amounts in Table 9 were calculated within the MReader Platform. The averages of the experimental class seem higher due primarily to one student, who read over 511,736 words. If this student were omitted, the experimental class average would be 203,618 words—slightly lower than that of the control group.

Table 9. *Total Words Read*

Class	Total (Mean)
Control Class	209,511
Experimental Class	230,334

Qualitative Investigation: Research Question 2

In order to obtain Personal Construct Repertory grids, 144 bipolar constructs were elicited from the learners. In order for graph readouts to be presented legibly, only numbered designations (e.g., *C6* meaning *Construct #6*) were entered into the analysis software packages. Figure 3 shows a graphical representation of the correlations between the elements and the constructs using Principal Components Analysis. The elements were *Activity I hate*, *Book reports*, *Activity I love*, *Vocabulary log*, and *Research matrix - corpus*. *Corpus* was used instead of *DDL* because it has entered the Japanese language primarily due to the work of Tono, who appeared on a nationally syndicated educational program on Japanese television and produced a number of teaching materials to supplement Japan's NHK educational programs (e.g., Tono, 2003, 2005). The specifics of liked and disliked activities will be mentioned shortly, but from the 144 elicited bipolar constructs, Figure 3 features the two which were most strongly rated by the experimental group, and which provided a correlational matrix for 80.2% of the data.

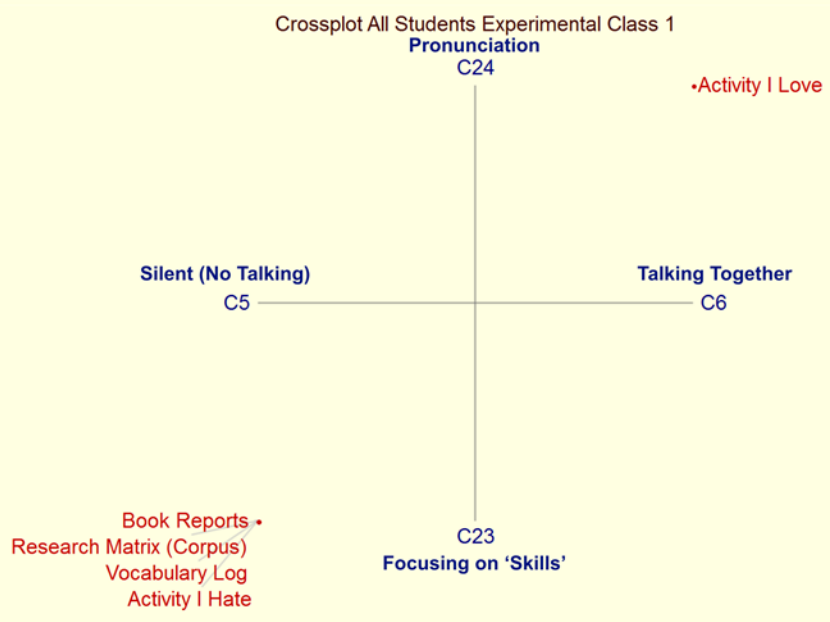


Figure 3. Most highly rated constructs for experimental group

These constructs were *Talking together* as opposed to *Silent (no talking)*, and *Pronunciation* as opposed to *Focusing on skills*. We confirmed later with this class that what they meant by pronunciation referred to learning how to emulate high status American or British accents—something that was outside the purview of a reading class. Nevertheless, when viewed two-dimensionally, it appears that virtually every element of the DDL class clustered between the negatively valued constructs. Flattening out the data in this manner, however, distorts relations much as a Mercator map distorts the globe, so a more nuanced view is gained by adding the third most strongly rated bipolar construct, that of *Useful and challenging* as opposed to *Boring and meaningless.* This reveals that the DDL activities were construed as almost identical to the vocabulary logs, and that while these activities were seen to make students silently focus on skills, DDL was also generally accepted as both useful and challenging (see Figure 4).

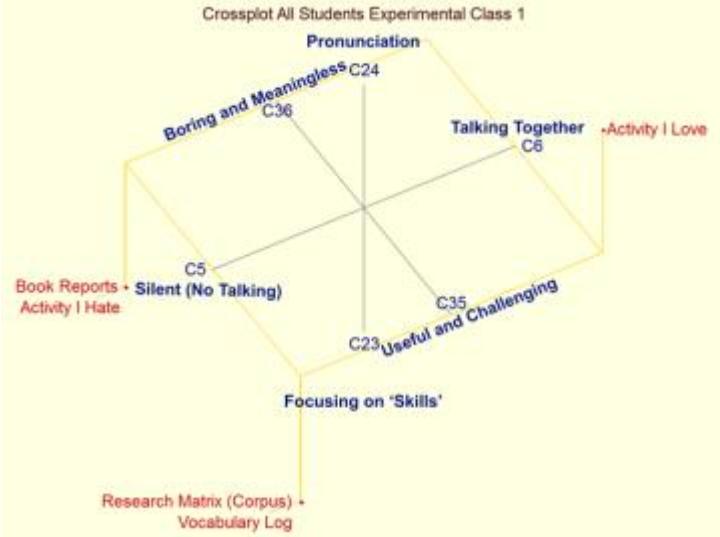


Figure 4. Multidimensional graph of experimental group constructs and elements

Figure 5 reveals a wider sample of the more heavily weighted constructs. The snapshot indicates that even as some students saw DDL as useful and challenging (C35), others felt that it was *simply searching for rules* (C96). With DDL, one *had to work hard to find an answer* (C46), but it was still felt as if one was *only studying words* (C55), or engaged in *rote vocabulary learning* (C58). Although seen as *mechanical* (C65), there was tacit recognition that it helped in *learning how to use new vocabulary* (C73). While students *had to write, though thankfully not a book report* (C75), and that it entailed *writing in isolation* (C99), something that was patently *boring* (C94), students also recognized it as *good for memorization* (C97), which in an East Asian context, is a term thought to be synonymous with learning.

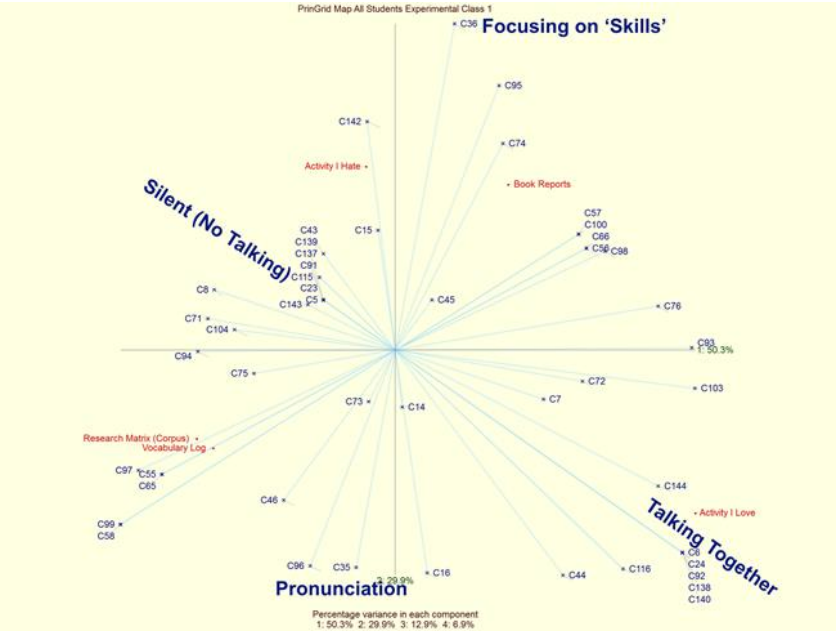


Figure 5. Constellation of highly rated constructs for the experimental group

The mixed response to a semi-hard approach to DDL is supported by the expressed likes and dislikes of this particular group. Individual elements of activities that students loved were those such as *reading books*, *pair work activities*, *singing songs*, *group work*, and *role-play*. These are the activities that elicited

constructs such as *talking together* (C6), *sharing opinions through conversations* (C44), *cooperative work with friends* (C92), and *expressing oneself through conversations* (C103). Bridging preferences were those such as *thinking in English* (C7), *reflecting upon new knowledge* (C14), *drawing pictures* (C16), and *learning pronunciation through singing* (C24). Other constructs in this quadrant's cluster emphasized *having the chance to just read* (C116), which was also expressed by other students as *focusing on reading* (C138), and *simply reading* (C144), because, as one student put it, *reading is what I love* (C140).

Practices that were strongly disliked were those of *writing compositions*, *grammar practice*, *reading in front of the class, giving presentations*, *practicing English with someone who speaks the same mother tongue*, and *reading textbooks or encyclopedias.* These clustered around constructs such as *silent (no talking)* (C5), *studying just for information* (C8), and *getting dry information from a book* (C43). They disliked *having to think about the unknown* (C15), *working on skills* (C23), and *having to do some sort of exercise after reading* (C139). This not only included *having to study word meanings* (C71), but also *writing book reports* (C137), *writing things out for people* (C115), *writing on paper* (C91), or do*ing activities that are both simple and boring* (C104), such as *having group discussions based on the book report* (C142) or other *discussion activities based on what one has read* (C143).

Discussion

It can be seen quantitatively that while the lexicogrammatical proficiency of both the experimental and control groups significantly improved over the semester, the control group showed greater improvement. The qualitative data clearly indicates resistance on the part of the experimental group to even a semi-hard approach to DDL. Even though it was seen as useful, the learners deemed it not worth the investment of time needed to make it work, preferring instead to focus on learning activities that are more in line with the classic ER approach—that is, simply reading and discussing their books with classmates.

This lack of student engagement with DDL raises the question of whether anything, either positive or negative, can be said about its effects in this study. However, it was ironically the control group, who was far more engaged in regularly keeping vocabulary logs, who were able to demonstrate greater improvement, suggesting that Ellis (2003, p. 573) may be correct in his assertion that any focused form of supplementary study is better than reading alone.

One of the stated aims of this study was to identify the conditions within which DDL might be contextualized for lower proficiency learners. This study finds that hard DDL (or even a semi-hard version of the approach using comprehensible corpus data) is inappropriate for young undergraduate learners at lower levels. DDL, as used by Johns (1991) at the University of Birmingham and by the second author at the University of Oxford, was carried out with graduate students, most likely to be in their late 20s and 30s. They were self-starters, worked on their own, and formulated questions as they developed hypotheses and carried out empirical investigations. Dispositions such as these fit well with hard DDL, because, in terms of data-driven learning, it is not only the data, but also the learners that drive the educational process. In such circumstances, the affective appeal of the materials is less important, and students are more willing to work with technology going through blocks of KWIC data to find answers to their questions.

Younger undergraduates in the ELT context, on the other hand, often expect more in the way of teacher direction, and tend to work from pre-packaged foreign language materials emphasizing fluency, group work, and self-discovery. Many have yet to be trained in inductive or deductive styles of learning. Therefore, we surmise that even if a corpus features language that is comprehensible to undergraduate learners, educational training in higher-level academic skills also seems to be a vital element in the potential success of a hard DDL approach. In addition, for learners at this stage of maturity and development, the affective element is key, as indicated by the qualitative data in this study. With regard to this, we feel the technology and unfamiliarity of DDL compounded the discomfort that students felt regarding the inclusion of the computer in the classroom. What we saw as a helpful tool, they construed as a space similar to one where they would be in front of the class, and where their questions or lack of skill in using the computer

might result in their losing face in front of their peers. Although we initially feared that the lack of online corpus access, as stipulated by OUP, would preclude adequate exposure to the Bookworms data and lead to student dissatisfaction, we now believe this factor is less likely to have had a major impact on their response to DDL. For these students, the most engaging learning activities involved group or pair work, interacting, sharing feelings and other emotive activities. Convenient access to the corpus would therefore not necessarily make DDL more appealing to such learners. Rather, it is necessary to take account of their clear learning preferences by, in this case, designing DDL materials which promote affective engagement.

From our observations of the improved responses to the inclusion of group-based DDL activities within an ER context, we believe that a repackaged, recontextualized form of soft DDL might be more effective, one that is data-directed rather than data-driven; one which is directed by the teacher for the learners in an explicit, structured way, and is as affectively engaging as possible. The impression of solitary hard study needs to be masked by purposeful activities that foster enjoyable social interaction in a non-threatening environment. Indeed, we have already been implementing this approach with students in our continuing investigation of this subject, and have found far more promising results that suggest that soft DDL is more effective with lower proficiency learners. These findings will soon be presented.

Most studies involving classroom research face shortcomings, and ours is no exception. While all of the analyses were undertaken in accordance with established research practices, the small sample sizes should be considered. The first thing that should be mentioned is that size alone does not necessarily undermine results that are statistically significant. The significant findings in this study indicate that something worth noticing has taken place. Nevertheless, the small number of participants also required us to approach our findings with humility, and with the hope that they might pave the way for further replication research. Also, while Nation and Beglar's (2007) VLT is a highly respected tool for exploring the lexical proficiency of English language learners, we suspect that that the lexical items do not fully represent the lexis contained within the Bookworms Graded Readers. We are creating a test instrument that follows the same structure and format as the VLT, but which uses vocabulary based on word frequencies derived from the Bookworms Corpus. We believe this will provide a far more sensitive and focused indicator of lexical learning in subsequent studies. Related to the lexical dynamics of this study, we suspect that the inclusion of vocabulary log sheets impeded a clearer snapshot of DDL's effects. In the studies that we will present later, we compare the use of soft DDL in an ER class without the use of vocabulary logs or other forms of explicit grammatical or lexical guidance. We believe this will allow for clearer investigation of the claims made both for ER free from explicit lexicogrammatical guidance and for the learning potential of DDL with lower proficiency learners.

Conclusion

In the search for greater insight into ways of improving second language instruction, discoveries of what does not work are as important as finding out what does work. Both represent progress. The findings of this study suggest that use of comprehensible semi-hard DDL in an ER program for lower proficiency learners did not lead to greater gains in reading speed, lexical knowledge, or grammatical improvement than those achieved by comparable students in a non-DDL ER program. Affectively, learners found semi-hard DDL to be useful but ultimately unattractive, preferring instead to focus on reading for pleasure and enjoying conversations of self-discovery with classmates. Based upon these findings, research is underway into ways of further softening DDL in an attempt to discover the conditions within which it might have the greatest impact for lower proficiency learners in the ELT context.

Acknowledgements

This work was supported in part by JSPS KAKENHI Grant Number JP16K2878. Appreciation is expressed to Oxford University Press for permission to use their materials in this research.

References

Alderson, J. (1979). The cloze procedure and proficiency in English as a second language. *TESOL Quarterly*, *13*(2), 219–226.

Allan, R. (2009). Can a graded reader corpus provide 'authentic' input? *ELT Journal*, *63*(1), 23–32. doi: 10.1093/elt/ccn011

AnalystSoft. (2016). StatPlus:mac Pro [Computer software]. Walnut, CA: AnalystSoft Inc.

Anthony, L. (2014). AntConc (Version 3.4.3m) [Computer software]. Tokyo, Japan: Faculty of Science and Engineering, Waseda University.

Bachman, L. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, *16*(1), 61–70.

Bamford, J., & Day, R. (2004). *Extensive reading activities for teaching language*. Cambridge, UK: Cambridge University Press.

Bannister, D., & Fransella, F. (1986). *Inquiring man: The psychology of personal constructs*. Kent, UK: Croom Helm.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15–38). Amsterdam, Netherlands: John Benjamins.

Bieri, T. (2015). ER definitions and practice: Instructor voices. *Extensive Reading in Japan*, *8*(1), 4–6.

Boulton, A. (2007). But where's the proof? The need for empirical evidence for data-driven learning. In M. Edwardes (Ed.), *BAAL annual conference 2007* (pp. 13–16). Edinburgh, UK: Scitsiugnil Press.

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, *35*(1), 81–106.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 1–46. doi: 10.1111/lang.12224

Bowen, J., Madsen, H., & Hilferty, A. (1985). *TESOL: Techniques and procedures*. Rowley, MA: Newbury House Publishers.

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, *19*(3), 307–328. doi: 10.1017/S0958344007000535

Chambers, A. (2010). What is data-driven learning? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 345–358). New York, NY: Routledge.

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, *6*(4), 289–302.

Cohen, L., Manion, L., & Morrison, K. (2003). *Research methods in education* (5th ed.). New York, NY: Routledge.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg, France: Language Policy Unit. Retrieved from http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. doi: 10.2307/3587951

Dao, T. N. (2014). Using internet resources for extensive reading in an EFL context. *Hawaii Pacific University TESOL Working Paper Series*, *12*, 72–95.

Day, R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, *14*, 136–141.

Day, R., & Bamford, J. (2004). *Extensive reading in the second language classroom*. New York, NY: Cambridge University Press.

Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. Long (Eds.), *The handbook of second language research* (pp. 256–310). Oxford, UK: Blackwell Publishing.

Easterby-Smith, M. (1981). The design, analysis, and interpretation of repertory grids. In M. L. G. Shaw (Ed.), *Recent advances in personal construct technology* (pp. 9–30). London, UK: Academic Press.

Edmonds, W. A., & Kennedy, T. (2017). *An applied guide to research designs: Quantitative, qualitative, and mixed methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Eichhorst, D., & Shearon, B. (2013). *ER@TU - 多読のすすめ The Tohoku University extensive reading manual*. Sendai, Japan: Tohoku University Higher Education Development Center.

Ellis, R. (2003). *The study of second language acquisition*. Oxford, UK: Oxford University Press.

Fromm, M., & Bacher, A. (2014). GridSuite 4 [Computer software]. Stuttgart, Germany: University of Stuttgart.

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ: The Electronic Journal for English as a Second Language*, *8*(4), 1–37.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). New York, NY: Routledge.

Gordani, Y. (2013). The effect of the integration of corpora in reading comprehension classrooms on English as a foreign language learners' vocabulary development. *Computer Assisted Language Learning*, *26*(5), 430–445.

Grabe, W., & Stoller, F. (2013). *Teaching and researching reading* (2nd ed.). New York, NY: Routledge.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer learner corpora, second language acquisition, and foreign language teaching*. Amsterdam, Netherlands: John Benjamins.

Greenwood, J. (1988). *Class readers*. Oxford, UK: Oxford University Press.

Hadley, G. (2001). Concordancing in Japanese TEFL: Unlocking the power of data-driven learning. In K. Gray, M. Ansell, S. Cardew, & M. Leedham (Eds.), *The Japanese learner: Context, culture, and classroom practice*. Oxford, UK: Oxford University, Department of Continuing Education.

Hadley, G. (2002). An introduction to data-driven learning. *RELC Journal*, *33*(2), 99–124.

Hadley, G., & Evans, C. (2001). Constructions across a cultural gap. In J. Edge (Ed.), *Action research* (pp. 129–143). Alexandria, VA: TESOL.

Hesse-Biber, S., & Johnson, R. (2015). *The Oxford handbook of multimethod and mixed methods research inquiry*. Oxford, UK: Oxford University Press.

Hirata, Y., & Hirata, Y. (2015). Data-driven learning and semi-structured interviews in tertiary language education in Japan. *International Journal of Innovation and Learning*, *18*(3), 313–332.

Jankowicz, D. (2004). *The easy guide to repertory grids*. Chichester, UK: Wiley.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning materials. *English Language Research Journal*, *4*, 1–16.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293–313). Cambridge, UK: Cambridge University Press.

Kelly, G. A. (1955). *A theory of personality*. New York, NY: Norton.

Kelly, G. A. (1963). *A theory of personality: The psychology of personal constructs*. New York, NY: Norton.

Kelly, L. (1969). *25 centuries of language teaching*. Rowley, MA: Newbury House.

Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-Test. *Language Testing*, *1*, 134–146.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *73*(4), 440–464. doi: 10.2307/326879

Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–24). New York, NY: Routledge.

Loschky, L. (1994). Comprehensible input and second language acquisition. *Studies in Second Language Acquisition*, *16*(3), 303–323. doi: 10.1017/S0272263100013103

McEnery, T., & Wilson, A. (1997). Teaching and language corpora (TALC). *ReCALL*, *9*(1), 5–14. doi: 10.1017/S0958344000004572

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge, and use: New perspectives on assessment and corpus analysis* (pp. 57–78). Colchester, UK: EuroSLA. Retrieved from http://www.eurosla.org/monographs/EM02/EM02tot.pdf

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, *22*, 1–18.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Plonsky, L., & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. doi: 10.1111/lang.12079

Quinn, E., Nation, I. S. P., & Millett, S. (2007). *Asian and Pacific speed readings for ESL learners: Twenty passages written at the one thousand word level*. Wellington, New Zealand: University of Wellington, English Language Institute.

Roberts, J. (1999). Personal construct psychology as a framework for research into teacher and learner thinking. *Language Teaching Research*, *3*(2), 117–144. doi: 10.1177/136216889900300203

Rowsell, L. (1992). Adults dropping out? Try Repgrid! *ELT Journal*, *46*(4), 362–372. doi: 10.1093/elt/46.4.362

Schlecty, P. (2011). *Engaging students: The next level of working on the work*. San Francisco, CA: Jossey-Bass.

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: Theoretical underpinnings and practical suggestions. *ELT Journal*, *49*(2), 133–143.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Shaw, M. L. G. (1980). *On becoming a personal scientist: Interactive computer elicitation of personal models of the world*. London, UK: Academic Press.

Shaw, M. L. G., & Gaines, B. (2010). Rep 5 (Version 1.04) [Computer software]. Cobble Hill, BC: Centre for Person-Computer Studies.

Shaw, M. L. G., & McKnight, C. (1981). *Think again: Personal problem-solving and decision-making*. Englewood Cliffs, NJ: Prentice-Hall.

Shimamoto, T. (2008). Anketo chosa kekka, daigaku [Results of university questionnaire]. In A. Yamazaki (Ed.), *Eigoryouku o takameru tame no tadoku jugyou no suishin [Extensive reading to improve learners' English language proficiency]* (pp. 54–65). Yokohama, Japan: Musashi Kogyo University.

St. John, E. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning & Technology*, *5*(3), 185–203. Retrieved from http://llt.msu.edu/vol5num3/stjohn/default.html

Sun, Y.-C., & Wang, L.-Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, *16*(1), 83–94. doi: 10.1076/call.16.1.83.15528

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. doi: 10.1016/0959-4752(94)90003-5

Tomlinson, B. (2016). Achieving a match between SLA theory and materials development. In B. Tomlinson (Ed.), *SLA research and materials development for language learning* (pp. 3–22). New York, NY: Routledge.

Tono, Y. (2003). *100-go de Start! Eikaiwa [Let's start English with 100 words!]* (USA ed.). Tokyo, Japan: NHK Publishing.

Tono, Y. (2005). *Supa Corpus Renshucho [Super corpus practice books]*. Tokyo, Japan: NHK Publishing.

Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 37–62). Amsterdam, Netherlands: John Benjamins.

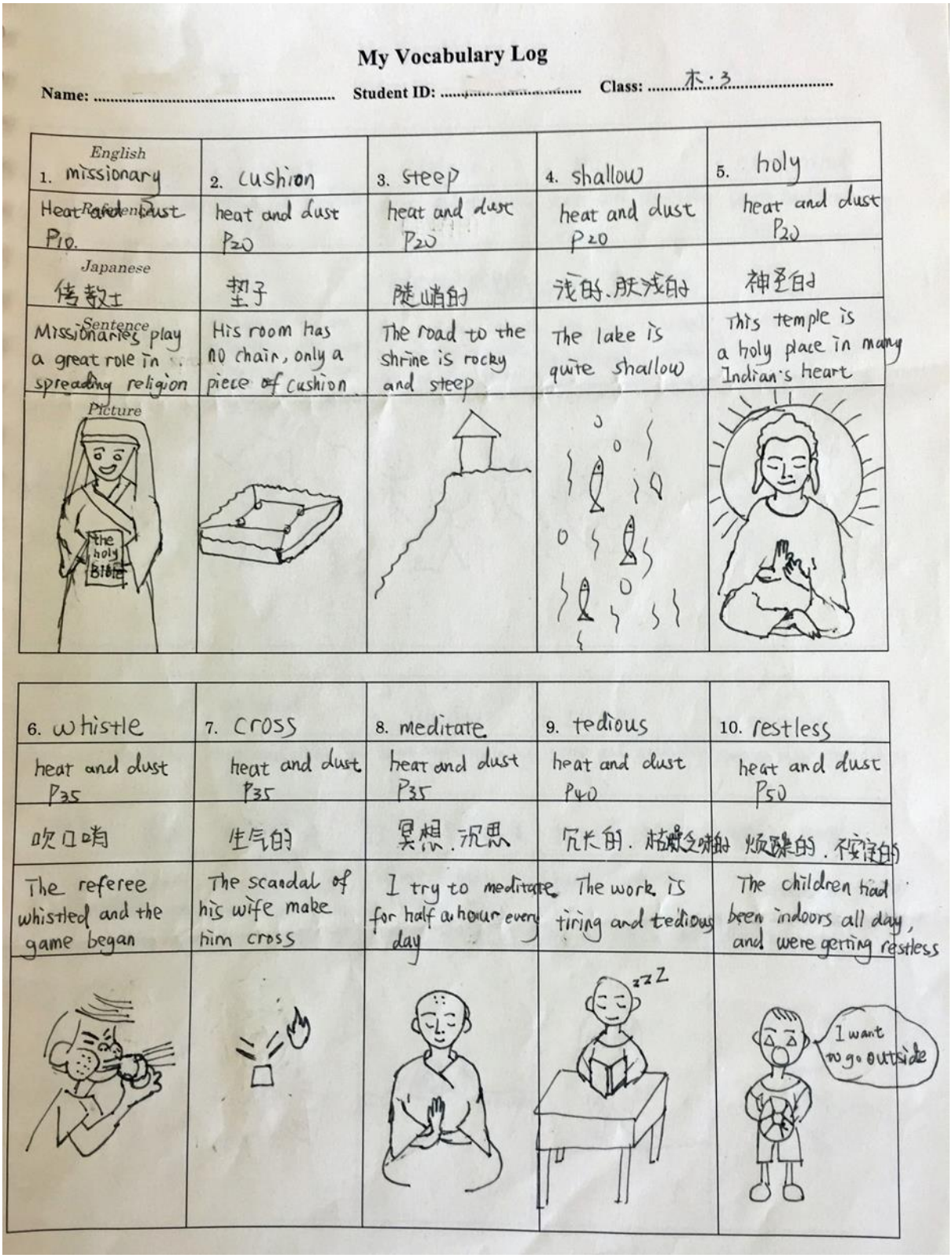
Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Waring, R., & McLean, S. (2015). Defining extensive reading. *Extensive Reading in Japan*, *8*(1), 7.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130–163.

Yoshida, H. (2014). An approach to extensive reading: Active involvement during sustained silent reading. *The Language Teacher*, *38*(6), 19–22.

Appendix A. Example Vocabulary Log



My Vocabulary Log

Name: ..... Student ID: ..... Class: 木・3

English	1. missionary	2. cushion	3. steep	4. shallow	5. holy
	Heat and dust P10.	heat and dust P20	heat and dust P20	heat and dust P20	heat and dust P20
Japanese	传教士	垫子	陡峭的	浅的、肤浅的	神圣的
Sentence	Missionaries play a great role in spreading religion	His room has no chair, only a piece of cushion	The road to the shrine is rocky and steep	The lake is quite shallow	This temple is a holy place in many Indian's heart
Picture					

	6. whistle	7. cross	8. meditate	9. tedious	10. restless
	heat and dust P35	heat and dust P35	heat and dust P35	heat and dust P40	heat and dust P50
	吹口哨	生气的	冥想、沉思	冗长的、枯燥乏味的	烦躁的、不安的
	The referee whistled and the game began	The scandal of his wife make him cross	I try to meditate for half a hour every day	The work is tiring and tedious	The children had been indoors all day, and were getting restless

Appendix B. Research Matrix

The Research Matrix

Question

Example: "How is _____ used properly?" "What is the difference between _____ and _____?"
"Why is _____ incorrect?" "What does _____ mean?"

Your Question: _____

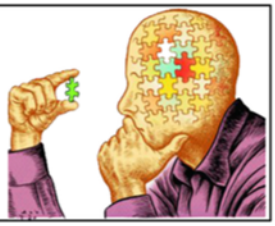
Investigate

Look at the corpus data. Do you see any regular patterns? Are there words that often appear before or afterwards? Is the a context or a situation where the word or words are used? Write down seven to eleven examples here:

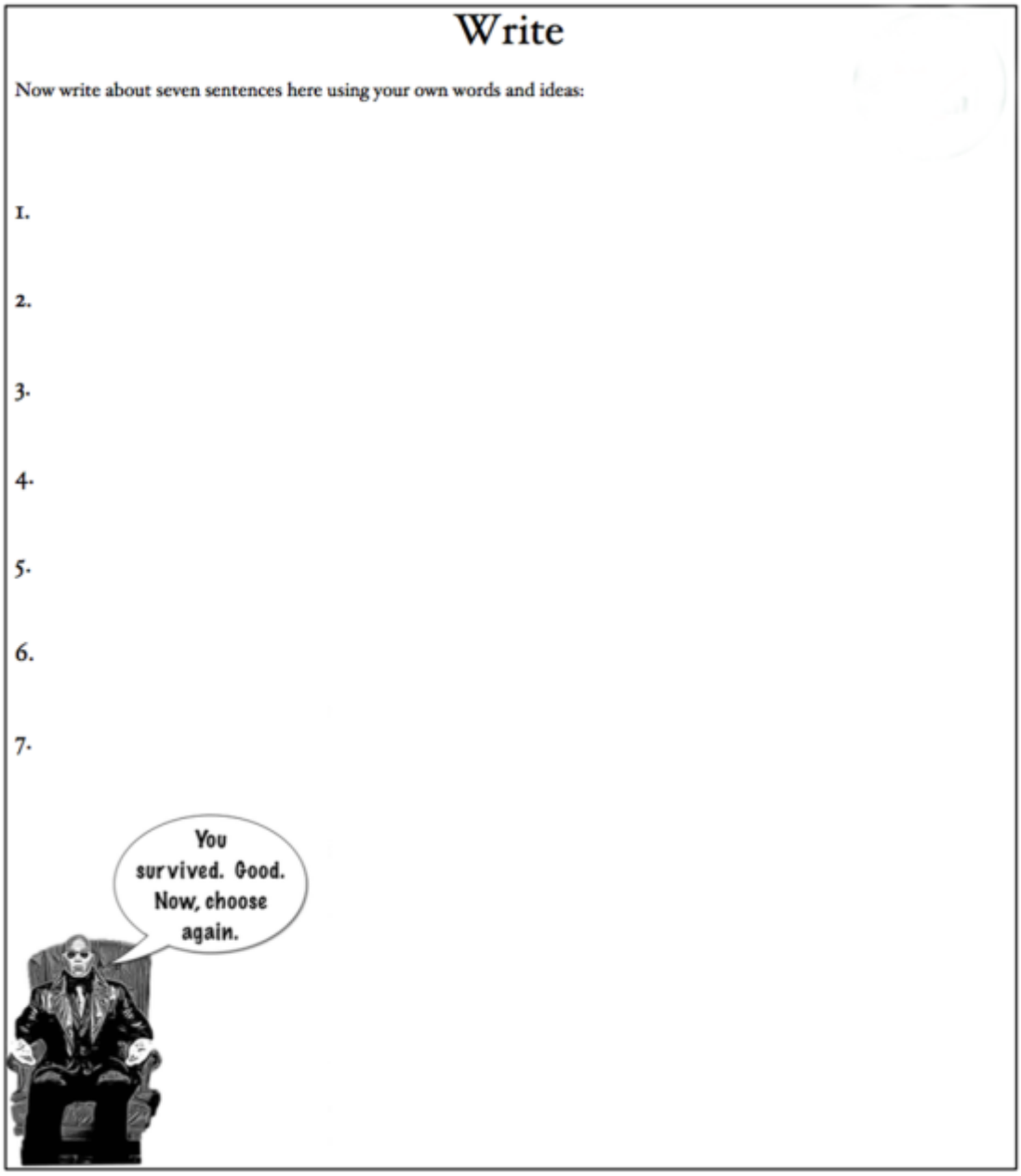
Studies show that students need to use a new word or phrase at least 7 to 11 times before they begin to learn and remember!

Think

Try now to write your own grammar rule for what you have studied:



Turn the Page Over



About the Authors

Gregory Hadley is a Professor of Applied Linguistics and Western Cultural Studies at Niigata University, and a Visiting Fellow at Oxford University. His latest book is *Grounded theory in applied linguistics research: A practical guide* (Routledge, 2017).

E-mail: ghadley@human.niigata-u.ac.jp

Maggie Charles is a Tutor in EAP at Oxford University Language Centre, where she teaches academic writing to graduates. She has published widely on academic discourse and corpus pedagogy. Recent books are *Introducing EAP* with Diane Pecorari (Routledge, 2015) and *Corpora, grammar, and discourse* with Nick Groom and Suganthi John (John Benjamins, 2015).

E-mail: maggie.charles@lang.ox.ac.uk